ELSEVIER
Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Engineering 15 (2011) 4178 – 4182
Procedia Engineering
www.elsevier.com/locate/procedia

Advanced in Control Engineering and Information Science

Scientific creativity research based on generalizability theory and BP_Adaboost RT

Jun Wang, Jiayuan Yu*

Department of Psychology, Nanjing Normal University, Nanjing,210097,China

Abstract

In this study, the scientific creativity of engineering students was measured. The quality of data was analyzed with Generalizability Theory. The modeling was conducted with BP_Adaboost RT, and compared with the model of multiple linear regression and single BP network. The results showed that Generalizability Theory could be applied to analyze the scientific creativity data. The quality of data would affect the predictive accuracy of the model. BP_Adaboost RT model was better than other two models.

Keywords: BP networks; Adaboost；data quality; Generalizability Theory; scientific creativity.

1. Introduce

Data mining has been applied in more and more areas. The importance of data quality has been recognized for more researchers. The effect of data quality to modeling error is also worth to study.

* Corresponding author. Tel.: +86-25-83598442; fax: +86-25-83598861
E-mail address: yujiayuanwx@163.com

doi:10.1016/j.proeng.2011.08.784

1.1. Data quality in scientific creativity assessment

The data of scientific creativity is usually obtained by self-rating scale and performance rating scale. It will be affected by subjective factors. Therefore the problem of data quality is remarkable.

In the history of psychological measurement, the assessment method of data quality has been explored for many years. It is showed in the three main psychological measurement theories.

In Classical Test Theory (CTT), the reliability is used to describe the test accuracy. In Item Response Theory (IRT), information function is used to describe the accuracy of measurement data. Because there are some shortcomings in CTT and IRT, the Generalizability Theory (GT) is more attractive for solving the data quality problem.

1.2. Generalizability Theory

In Generalizability Theory [1], universe score is used to represent level of the examinee's latent trait. The factors which affect the assessment scores are called as facets. In scientific creativity measurement, examinee's scores will be affected by items and judges. If the amount of items or judges are changed, the accuracy of the test scores will be changed too.

In Generalizability Theory, the measurement errors are divided into two categories: relative error and absolute error. Generalizability coefficient $E\rho^2$ is used to describe the relative error. Reliability index Φ is used to describe absolute error.

In this research, scientific creativity of the engineering students has been measured. The data quality and its affect on modeling have been explored.

2. Creativity measurement

2.1. Scientific creativity

Scientific creativity is the ability in learning scientific knowledge and solving scientific problem. It is a part of creativity and is very important for engineering students

Adolescent Scientific Creativity Scale (ASCS) developed by Weiping Hu is often used to measure the scientific creativity [2]. It could measure the seven aspects of scientific creativity, including object use, problem giving, product improvement, science fancy, problem solving, science experiment and product design .The total score of above seven aspects represents the level of scientific creativity.

2.2. Creative affective

Recent research has showed that creativity is not only related to thinking but also affective factors. Williams defined creativity in relation to four cognitive factors (fluency, flexibility, originality, and elaboration) and four affective factors (risk-taking, curiosity, imagination, and complexity). Williams Creativity Test B (WCTB) [3] was used to measure the creative affective factors. This is a 50-item creativity assessment instrument that provides scores for risk taking, curiosity, imagination and complexity.

In this research, the creative affective and scientific creativity were studied synthetically. The creativity model was set up for engineering students. The creative affective factors were used to predict the scientific creativity.

3. Data collection

3.1. Subject

In two universities at Nanjing, engineer students of mechanics, computer, chemistry and architecture were selected randomly. 780 students' data were analyzed after the ineffective questionnaires were deleted.

3.2. Measuring and Scoring

The creativity test was conducted with each class as a group. For avoid order error, one class was conducted WCTB at first, another class in the same specialty was conducted ASCS at first. The test time of WCTB was 25 minutes, and the test time of ASCS was 60 minutes. Three judges rated the examinees' response to ASCS respectively.

4. Data Analysis

4.1. Generalizability theory based data analysis

In the scientific creativity measurement, the object was creativity scores. Item and judge were the factors which affected the test scores, they were facets.

Because every examinee answered every item, every judge rated every examinee's response, it became a random two-facet cross design ($p\times i\times r$).

There were two steps in data analysis with generalizability theory. G study estimated the different error sources and variance of cross effect. D study estimated the relative error and absolute error, also computed the generalizability coefficient and reliability index.

In this research, generalizability coefficient $E\rho^2$ and reliability index Φ of scientific creativity data were obtained under the numbers of judges were 1, 2 and 3 respectively. The results were in the Table 1.

Table 1 Generalizability coefficient and reliability index under different numbers of judges

n	1	2	3
$E\rho^2$	0.8257	0.8639	0.8976
Φ	0.7923	0.8136	0.8461

It showed that the reliability of data was the lowest when there was only one judge. The reliability improved along with the increasing of the numbers of judges. However, the errors were still obvious because of the subjective assessment.

4.2. BP_Adaboost RT algorithm

BP neural networks were often applied to set up the model for describing the relationship of variables in psychology. Due to the BP networks used the grads ascent algorithm, it could not ensure to get the universal minimum. Therefore it could only be weak predictor. To overcome this difficulty, several methods which combined some BP networks into a stronger predictor had been promoted. One of these ensemble learning algorithms was Boosting algorithm. Adaboost algorithm which was easier to apply was promoted based on Boosting by Freund in 1995. Solomatine and Shrestha improved this method, and promoted Adaboost RT algorithm in 2004 [4]. They introduced threshold Φ, compared the train error with threshold, and divided the train set into good and bad categories. The algorithm pays more attention to the learning difficult data. It could give iterative times arbitrarily. The last output would be the weighted average of all weak predictors.

BP_Adaboost RT algorithm used the BP networks as weak predictors, and several BP networks was combined. Following were steps of BP_Adaboost RT algorithm [5]:

(1) Input: make train set, define weak predictor, maximum iterative times and threshold Φ.

(2) Initialize: Let weight distribution is $D_t(i)=1/m$, when $t=1$, m is the number of train sample. Let initial error is zero.

(3) Iterative: Training BP networks and setting up regression model $f_t(x)$. Computing the train error:

$$ARE_t(i)=\left|\frac{f_t(x_i)-y_i}{y_i}\right|$$

Update weight D_t:

$$D_{t+1}(i)=\frac{D_t(i)}{Z_t}\times\begin{cases}\beta_t & ARE_t(i)\le\phi\\ 1 & others\end{cases}$$

Z_t was standardization factor in above formula.

4.3. Modeling with BP_Adaboost RT

(1) Research design

In this research, the creative affective scores of engineering students were the input of the model, scientific creativity scores were the output of the model. Scientific creativity scores were obtained under three conditions, there were 1, 2 and 3 judges respectively. The data quality of each condition was different.

In order to compare the BP_Adaboost RT method with single BP network and multivariable linear regression, three modeling were set up respectively. Therefore the 3X3 research design was made. Following was the modeling steps with BP_Adaboost RT.

(2) Data selection and network initial

In this research, the students were divided into two parts, which were 70% and 30% respectively, i.e. 546 and 234 students. They were used as training sample and test sample respectively.

BP networks were weak predictors in this study. The dimensions of the network input and output were dependent on the sample. The four aspects of creative affective (risk taking, curiosity, imagination and complexity) were the input of BP networks. The scientific creativity was the output of BP networks. The number of hidden neurons were 2. So that the structure of network was 4-2-1.

(3) Wear predictor training

10 BP networks were trained. The parameters were as following: Maximum learning times was 2000, learning speed was 0.8, learning accuracy was 0.1.

(4) Update weight D_t

(5) Obtaining stronger learning function

The maximum training time was 5. The last predictor function was the integration of BP networks which were trained after 5 times.

(6) Computing the error of predictive model

The creative affective scores of test sample were input to the integrative BP networks which had been trained, and the real output $\hat{y}_i$ was obtained. The difference between real output y_i and expect output was computed with RMSE.

(7) Comparing the BP_Adaboost RT modeling and multivariable regression modeling, single BP network modeling

Based on above research design, the BP_Adaboost RT, multivariable regression and single BP network modeling were conducted. The RMSE of real output and expect output were computed, and were compared among these three models.

Because the scientific creativity scores were obtained with subjective assessment, the data quality were different when the distinct numbers judges were used. In above three modeling processes, the models from different data root in dissimilar numbers of judges were compared too. The results were in Table 2.

Table 2 The RMSE from different models

Numbers of judges	1	2	3
Multivariable regression	0.2351	0.1863	0.1394
Single BP network	0.1753	0.1231	0.0854
BP_Adaboost RT	0.1218	0.0851	0.0379

From table 2, it could find the predictor errors were affected by data quality. When there was only one judge, the data quality was the lowest (as in Table 1 the generalizability coefficient and reliability index were the smallest), and the correspondent RMSE was the largest. The data quality was improved when the numbers of judges increased, and the RMSE was decreased. For the same quality data, the RMSE of BP_Adaboosting RT model was the smallest, and the RMSE of multivariable regression model was the largest.

5. Conclusion

Above research showed:

(1) For engineering students, the data quality of scientific creativity could be checked with generalizability theory. The data quality could be improved with increasing the numbers of judges.

(2) BP_Adaboost RT algorithm was the best one compared with single BP network and multivariable regression when the scientific creativity model was set up.

References

[1] Yang Z., Zhang L. *Generalizability theory of assessment and its application*. Beijing: Education Science Press; 2003.

[2] Hu W. *Adolescent scientific creativity development and training*. Beijing: Beijing Normal University Press; 2003.

[3] Duan J. The experiment research of student creativity training. *Journal of Chinese Education*. 2001; **6**: 24-28.

[4] Solomatine D P, Shrestha D L. A boosting algorithm for regression problem. In: *Proc. of the International Journal Joint Conference on Neural Networks*. Piscataway; 2004, p.1163-1168.

[5] Mao Z. Tian H. Wang Y. Molten steel end temperature soft sensing in LF based on hybrid model using AdaBoost,. *Chinese Journal of Scientific Instrument*. 2008; **29**: 662-667.